AD-A032 496 DRC INVENTORY RESEARCH OFFICE PHILADELPHIA PA F/G 12/1
KALMAN AND MOVING AVERAGE FILTERS FOR FORECASTING: SYSTEMATICS --ETC(U)
OCT 76 D A ORR
UNCLASSIFIED IRO-TR-2-77 NL

1 OF 1
AD A032496

END
DATE
FILMED
1 -77

AD-
TECHNICAL REPORT FC
~~TR-77-2~~
TR-2-77
see not inside cover

KALMAN & MOVING AVERAGE FILTERS FOR FORECASTING: SYSTEMATICS OF DEMAND PROCESSES AND EXTENSIONS

DRC INVENTORY RESEARCH OFFICE

October 1976

ROOM 800
U.S. CUSTOM HOUSE
2nd and Chestnut Streets
Philadelphia Pa. 19106

TCD 14 : Per Telecon w/
DRC IRO, 8/348-6984,
This Rept no. should be
TR-2-77, as was
TR-1-77, AD-A029715,

1 Dec 76

KALMAN & MOVING AVERAGE FILTERS FOR FORECASTING:
SYSTEMATICS OF DEMAND PROCESSES & EXTENSIONS

TECHNICAL REPORT

BY

DONALD A. ORR

OCTOBER 1976

DRC INVENTORY RESEARCH OFFICE
US ARMY LOGISTICS MANAGEMENT CENTER
ROOM 800 US CUSTOM HOUSE
2ND AND CHESTNUT STREETS
PHILADELPHIA, PA. 19106

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)

6 KALMAN ~~&~~ and MOVING AVERAGE FILTERS FOR FORECASTING: SYSTEMATICS OF DEMAND PROCESSES and ~~&~~ EXTENSIONS.

5. TYPE OF REPORT & PERIOD COVERED

9 Technical Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)

10 Donald A. Orr

8. CONTRACT OR GRANT NUMBER(s)

14 IRO-TR-2-77

9. PERFORMING ORGANIZATION NAME AND ADDRESS

DRC Inventory Research Office

US Army Logistics Management Center

Room 800, US Custom House, Phila., Pa. 19106

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

12 60 p.

11. CONTROLLING OFFICE NAME AND ADDRESS

US Army Materiel Development & Readiness Command

5001 Eisenhower Avenue

Alexandria, Va. 22333

12. REPORT DATE

11 Oct[redacted]76

13. NUMBER OF PAGES

59

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)

UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

Information and data contained in this document are based on input available at the time of preparation. Because the results may be subject to change, this document should not be construed to represent the official position of the US Army Materiel Development & Readiness Command unless so stated.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Forecast Techniques

Demand & Ancillary Processes

Kalman Filters

Weighted Moving Averages

Bayes, Box-Jenkins, Spectral Techniques

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This technical report is a digest of theoretical (mostly) results on short term forecasting. The ubiquitous criterion is mean square error MSE, which is the most amenable to mathematical analysis, but in some cases transformations of process variables to be forecasted infer other error measures. This study is also a proselytization for the Kalman filter, which together with its underlying vector process model, is a powerful and flexible algorithm which encompasses many other techniques. → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 403 572

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

mt

To an extent, the study concerns systematics, a scheme for classifying such well known forecasting algorithms as exponential smoothing, moving averages, regression line fitting, by the underlying processes for which these algorithms are optimal or sub-optimal.

If the process D of interest is dependent upon an ancillary process, a variable h of which is observable, the intuitive notion of using an estimate of a rate variable D/h for forecasting D - if h can be forecasted - is shown in mathematical terms to improve performance under some error measure. Kalman and weighted moving average filters on the rate or on a function of rate are found to be optimal and suboptimal respectively.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

This technical report is a digest of theoretical (mostly) results on short term forecasting. The ubiquitous criterion is mean square error MSE, which is the most amenable to mathematical analysis, but in some cases transformations of process variables to be forecasted infer other error measures. This study is also a proselytization for the Kalman filter, which together with its underlying vector process model, is a powerful and flexible algorithm which encompasses many other techniques. Most of the results are original, albeit many are first order derivations; i.e., anyone with the starting relations culled from various sources, could generate and manipulate the subsequent equations. A key point, however, is that an extensive synthesis has not been made previously in the literature; for example, many statisticians are unaware of Kalman's work and control theoreticians and engineers probably were not cognizant of Box-Jenkins work (publication of their classic book improved matters). To an extent, the study concerns systematics, a scheme for classifying such well known forecasting algorithms as exponential smoothing, moving averages, regression line fitting, by the underlying processes for which these algorithms are optimal or sub-optimal.

The most useful and most discussed process model in this study because of simplicity, flexibility, and robustness is the Dynamic Mean. (Process mean is a random walk and observations are corrupted by noise.) A first-order Kalman filter is the optimal algorithm and can handle prior distributions on the process mean and time varying variances of noise. After infinite observation time on the process ("steady-state"), the Kalman filter acts like single exponential smoothing if the noise variances are constant. In steady state, a moving average is a sub-optimal* algorithm for the Dynamic Mean. Other useful descriptive models of a process often

*In our context, sub-optimal refers to methods which track generated data of this model better than other models and which use the process parameters in determining algorithm parameters, e.g., moving average base period.

may be changed to a Dynamic Mean by a transformation of the observed process variable. This transformed variable, if observable, can then be forecasted and then an inverse transformation of this forecast will in some sense be a "best" forecast of the original process variable.

If the process D of interest is dependent upon an ancillary process, a variable h of which is observable, the intuitive notion of using an estimate of a rate variable D/h for forecasting D - if h can be forecasted - is shown in mathematical terms to improve performance under some error measure. Kalman and weighted moving average filters on the rate or on a function of rate are found to be optimal and suboptimal respectively. If the process D is dependent more indirectly on a secondary ancillary variable g, which may be "more" observable than h, it may be more efficacious to operate on the rate variable D/g. Relations among parameters of the D, D/h, and D/g processes and among parameters of the subsequent algorithms are found.

Expressions have been developed for the mean-square error when moving averages are applied to the observations from the process models.

The emphasis of the mathematical details is on scalar representations of the Kalman filter as applied to Dynamic Mean process models (no trend) and Linear Growth process models (trend) - with and without single ancillary variables. Operating from various initial conditions the Kalman filter can encompass algorithms obtained from exact or approximate Bayesian methods, linear regressions over time or the ancillary variable, and weighted moving averages. In steady-state the Kalman equations are equivalent to Box-Jenkins algorithms for integrated moving average processes (0,1,1) and 0,2,2), and to Weiner-Hopf filters in the frequency domain (spectral analysis).

Future work of interest includes demonstrating that the forecasting techniques for the Box-Jenkins general ARIMA model can be subsumed by the general vector Kalman filter. Also desirable are comparisons of performance when forecasting from a rate variable, say D/h, when the ancillary process h can be predicted with limited accuracy, versus forecasting directly from D. And there is need for detailed relations of error measures, for various transformations of the process, to the MSE measure.

PART I

BASIC VARIABLES, FOUNDATIONS

CHAPTER II

DYNAMIC MEAN PROCESS

2.1 Equations, Algorithms, Relations

The following equations represent the Dynamic Mean model.

$$\left.\begin{aligned} y_n &= x_n + \eta_n \\ x_n &= x_{n-1} + v_n \end{aligned}\right\} \quad (2.1)$$

$$\left.\begin{aligned} &E(\eta_n) = E(v_n) = 0 \\ &\mathrm{Var}\ (\eta_n) = r_n^2 \\ &\mathrm{Var}\ (v_n) = q_n^2 \\ &E(\eta_i \eta_j) = E(v_i v_j) = 0 \qquad j \neq i \\ &E(\eta_i v_i) = 0 \end{aligned}\right\} \quad (2.2)$$

y_n = observed value of process at time n

x_n = mean of process at time n

η_n = additive random noise

v_n = additive random change in mean x

This model is sufficiently complex to explain short term trends in a time series; the random walk process on x can generate these trends. The initial value $m_o = E(x_o)$ determines the long term process level. Moving averages and single exponential smoothing work well on this process.

Kalman Filter - 1st Order

This algorithm is optimal (see below) for the dynamic mean process. The "confidence" placed on the most recent observation y_n is expressed by the value of the weight G_n, which is a function of the noise variances.

$$\hat{y}_n(\ell) = \hat{x}_n \tag{2.3}$$

$$\hat{x}_n = \hat{x}_{n-1} + G_n \cdot (y_n - \hat{x}_{n-1}) \tag{2.4}$$

$$G_{n+1} = \frac{q_n^2 + r_n^2 G_n}{q_n^2 + r_n^2 G_n + r_{n+1}^2} \tag{2.5}$$

where

$\hat{x}_n$ = estimate of mean of process at end of period n

$\hat{y}_n(\ell)$ = forecast at end of period n of the process value ℓ periods later

G_n = variable weighting of one-step ahead error

Initialization:

$$\hat{x}_o = \mu + G_o(y_o - \mu) \tag{2.6}$$

$$G_o = \frac{\tau^2}{\tau^2 + r_o^2} \tag{2.7}$$

where

y_o = initial observed value of process

μ, τ^2 = mean and variance of a prior distribution on x_o

r_o^2 = variance associated with y_o

The Kalman filter minimizes the mean square error (MSE) in one-step ahead forecasts, i.e.,

$$E(y_n - \hat{y}_{n-1}(1))^2 \equiv V_{1,n}$$

and the estimator $\hat{x}_n$ is actually $E(x_n | \underset{\sim}{y}_n)$ where $\underset{\sim}{y}_n$ designates all past history on y thru period n. If the distributions are Gaussian, x_n is also the maximum likelihood estimator.

Also note

$$\mathrm{Var}\,(x_n | \underset{\sim}{y}_n) \equiv \sigma_n^2 = \frac{r_n^2\,(q_{n-1}^2 + \sigma_{n-1}^2)}{q_{n-1}^2 + \sigma_{n-1}^2 + r_n^2} = r_n^2\, G_n \qquad (2.8)$$

$$V_{1,n} = \mathrm{Var}(x_{n-1} | \underset{\sim}{y}_{n-1}) + \mathrm{Var}(x_n | x_{n-1}) + \mathrm{Var}(y_n | x_n)$$

$$= \sigma_{n-1}^2 + q_n^2 + r_n^2 \qquad (2.9)$$

(2.9 is obtained from

$$E[y_n - E(x_{n-1} | \underset{\sim}{y}_{n-1})]^2$$

$$= E[(y_n - x_n) + (x_n - x_{n-1}) + (x_{n-1} - E(x_{n-1} | \underset{\sim}{y}_{n-1}))]^2$$

and may be viewed by the usual breakout of MSE of a forecast:

r_n^2 = variance of process

σ_{n-1}^2 = variance of forecast

q_n^2 = bias squared

Optimal Exponential Smoothing - Steady State Process

Let the noise variances be constant,

$$r_n^2 = r^2 \ , \ q_n^2 = q^2 \qquad (2.10)$$

and let the process and the forecast thereof be far from the origin. Then Harrison [5] shows that the following exponential smoothing type algorithm minimizes $V_{1,n} = V_1$

$$\hat{y}_n(\ell) = \hat{x}_n$$

$$\hat{x}_n = \hat{x}_{n-1} + G_\infty (y_n - \hat{x}_{n-1}) = \hat{x}_{n-1} + G_\infty e_n \qquad (2.11)$$

$$G_\infty = (\sqrt{1 + 4k} - 1) / 2k \qquad (2.12)$$

$$k = r^2/q^2 \qquad (2.13)$$

and $$V_1^* = r^2 / (1-G_\infty) \qquad (2.14)$$

For a non-optimal G,

$$V_1 = (2 G r^2 + q^2) / (G (2-G)) \qquad (2.15)$$

$$C \equiv E(e_n e_{n+1}) = (1-G) V_1 - r^2 \qquad (2.16)$$

Also defining V_L as the MSE of the cumulative L - step forecast,

$$V_L = E\left[\sum_{\ell=1}^{L} (y_{n-1+\ell} - \hat{y}_{n-1}(\ell))\right]^2 \qquad (2.17)$$

then for optimal G

$$\left.\begin{aligned} V_L &= L \left[1 + (L-1) G_\infty + \frac{(L-1)(2L-1)G_\infty^2}{6}\right] V_1^* \\ &= V_1^* \Phi \ , \end{aligned}\right\} \qquad (2.18)$$

Φ being defined by the equation, and for non-optimal G,

$$\left.\begin{aligned} V_L &= \Phi V_1 + C \cdot 2\,\Psi \\ &\text{where } \Psi = [(L-1)G(L-1)^2 + G^2 \frac{L(L-1)(L-2)}{3}] \end{aligned}\right\} \qquad (2.19)$$

All equations except (2.19)[1] can be found in Harrison [5].

Now note that (2.5), given (2.10), in steady-state becomes

$$G_\infty = \frac{q^2 + r^2\, G_\infty}{q^2 + r^2\,(G_\infty + 1)} \qquad (2.20)$$

The solution of (2.20) is exactly (2.12).

Also note in steady state (2.9) becomes

$$V_1 = \sigma_\infty^2 + q^2 + r^2 = r^2\, G_\infty + q^2 + r^2 \qquad (2.21)$$

and (2.21) equals $r^2/(1-G_\infty)$, i.e., (2.14), only when G_∞ is given by (2.12). Equation (2.21) and its solution for V_1 relates the Kalman expressions for variances of error in steady state to the Harrison formulae.

<u>Moving Average Algorithm</u>

The algorithm is described by Equation (2.22)

$$\begin{aligned} \hat{y}_n(\ell) &= \hat{x}_n \\ \hat{x}_n &= \frac{1}{B} \sum_{j=1}^{B} y_{n-j+1} \end{aligned} \qquad (2.22)$$

Now under conditions (2.12) of constant variances, we show in Appendix B that

$$V_1 = q^2\,[k + k/B + \frac{(B+1)(2B+1)}{6B}] \qquad (2.23)$$

---

[1] Equation (6.18) in [5] is summed over L steps and squared, where only terms involving $E(e_j e_j)$, $E(e_j e_{j+1})$ are non-zero.

$$V_L = L^2V_1 - q^2[(L-1)Lk - \frac{(L-1)}{6}(2L^2-L)] \quad (2.24)^2$$

Minimizing V_1 under continuity assumptions on the moving average (MA) base B,

$$\frac{dV}{dB} = q^2\left[\frac{-k}{B^2} + \frac{4B+3}{6B} - \frac{(2B^2+3B+1)}{6B^2}\right] = 0$$

$$= \frac{q^2}{6B^2}\left[-6k + 4B^2 + 3B - 2B^2 - 3B-1\right] = 0 \quad (2.25)$$

From (2.25)

$$B^* = \sqrt{\frac{1+6k}{2}} \quad (2.26)$$

Conversely the k for which B^* is suboptimal is $(2B^2-1)/6$.

Equations (2.26) and (2.12) are important in that if k is known for the dynamic mean process, the best moving average algorithm uses a base B^* and the best exponential smoothing algorithm uses a smoothing constant G_∞. The choice of these parameters is not arbitrary and are certainly not related by Brown's [2] expression $G \sim 2/B+1$.

2.2 Quantifications

The first table compares the smoothing parameters or weights obtained using (2.12) and those obtained using Brown's relation which equates "average age" of data under moving average and exponential smoothing procedures.

[2] An important special case, for $q^2 = 0$, $k = \infty$, gives $V_L = L^2r^2/B + Lr^2$

TABLE 2.1: COMPARISON OF SMOOTHING PARAMETERS

B	$k = \frac{2B^2-1}{6}$	G	$G = \frac{2}{B+1}$
	0	1	
1	.7667	.8730	1
2	1.1667	.5916	.6667
4	5.1667	.3537	.4
8	21.1667	.1950	.2222
12	47.8333	.1345	.1538
∞	∞	0	0

Table 2.2 compares mean square forecast errors V_1, V_L using optimal weights G , V_1(opt); using MA base B, V(MAB); using G = 2/B+1, V(ESB). For comparison across different k's, $r^2 + q^2$ was held constant = 48.833, i.e., $q^2(1+k) = 48.833$ or $q^2 = 1$ when k = 47.833.

TABLE 2.2: COMPARISON OF MEAN SQUARE ERRORS

k	B^*	Leadtime L	V_1(opt)	V(MA12)	V(ES12)	V(MA8)	V(ES8)
47.8333	12	1	55.267	56.333	55.340	56.999	56.343
47.8333	12	4	323.986	341.332	332.652	351.998	371.304
21.1667	8	1	57.925	60.460	58.273	59.480	58.044
21.1667	8	4	397.026	438.638	384.467	422.973	410.927

As k increases - the process becoming more stationary - the V's decrease; it is easier to forecast a process where the random change (q^2 effect) in the mean is low. Across the rows the results are what one might expect, with one exception; exponential smoothing with G = 2/12+1 = .1538 does better forecasting over 4 periods than with G_∞ = .1950 obtained by minimizing MSE of 1 period forecasts for k = 21.1667. Reason: even though x_n obtained after period n from (2.11) minimizes $E(y_{n+1} - \hat{x}_n)^2$,

$E(y_{n+2} - \hat{x}_n)^2, \ldots, E(y_{n+L} - \hat{x}_n)^2$; it does not necessarily minimize, for $L > 1$,

$$E\left(\sum_{\ell=1}^{L} (y_{n+\ell} - \hat{x}_n)\right)^2$$

because of non-zero cross product terms. G = .1538 is more conservative in weighting recent observations and the algorithm thereby benefits since the mean may shift back to past values in the upcoming four periods.

Finally we note there is not a great degradation from using sub-optimal algorithms. Specifically, for L = 1,

k	21.1667	47.8333
$100[V(MAB^*) - V_1(opt)]/V_1(opt)$	2.68%	1.93%

The percentages increase for decreasing k; for k ~ 1 the degradation in performance is ~ 5%.

CHAPTER III

MODIFICATIONS

3.1 Linear Growth Model

We augment the dynamic mean process with a growth term β and random variation δ

$$y_n = x_n + \eta_n$$

$$x_n = x_{n-1} + \beta_n + v_n \tag{3.1}$$

$$\beta_n = \beta_{n-1} + \delta_n$$

$$E(\delta_n) = E(\delta_n \delta_m) = 0 \quad Var(\delta_n) = p_n^2 \tag{3.2}$$

and as before $Var\ \eta_n = r_n^2$, $Var\ v_n = q_n^2$

This model allows for linear growth over time of the process mean; the rate is more apparent if $E(\beta_o) >> p$. Double exponential smoothing and linear regression with time as the independent variable would do well.

Kalman Filter - 2nd Order

The optimal Kalman filter for the linear growth model is given by Equations (3.7) - (3.17).

$$\hat{y}_n(\ell) = \hat{x}_n + \ell \hat{\beta}_n \tag{3.3}$$

$$\hat{x}_n = \hat{x}_{n-1} + \hat{\beta}_{n-1} + G_n \cdot e_n \tag{3.4}$$

$$\hat{\beta}_n = \hat{\beta}_{n-1} + H_n \cdot e_n \tag{3.5}$$

$$e_n = y_n - \hat{y}_{n-1}(1) = y_n - (\hat{x}_{n-1} + \hat{\beta}_{n-1}) \tag{3.6}$$

$$V_n = s_{11,n} + r_n^2 \tag{3.7}$$

$$G_n = s_{11,n} / V_n \tag{3.8}$$

$$H_n = s_{12,n} / V_n \tag{3.9}$$

$$s_{11,n+1} = s_{11,n} - G_n^2 V_n + 2s_{12,n} - 2G_n H_n V_n$$

$$+ s_{22,n} - H_n^2 V_n + q_n^2 + p_n^2 \tag{3.10}$$

$$s_{12,n+1} = s_{12,n} - G_n H_n V_n + s_{22,n} - H_n^2 V_n + p_n^2 \tag{3.11}$$

$$s_{22,n+1} = s_{22,n} - H_n^2 V_n + p_n^2 \tag{3.12}$$

where

$$\begin{bmatrix} s_{11,n} & s_{12,n} \\ s_{12,n} & s_{22,n} \end{bmatrix} = \mathrm{Cov}\left(\begin{matrix} x_n \\ \beta_n \end{matrix} \,\middle|\, \underset{\sim}{y}_{n-1} \right) \tag{3.13}$$

Initialization:

$$\hat{x}_o = \mu + G_o(y_o - \mu) \tag{3.14}$$

$$\hat{\beta}_o = b \tag{3.15}$$

$$G_o = \tau^2 / (\tau^2 + r_o^2) \tag{3.16}$$

$$s_{11,o} = \tau^2 \; , \; s_{22,o} = \tau_\beta^2 \; , \; s_{12,o} = o \tag{3.17}$$

where b, τ_β^2 = mean and variance of a prior distribution on β_o

<u>Optimal "Double" Exponential Smoothing</u> - Steady State Process

As in Chapter II,

$$r_n^2 = r^2 \;,\; q_n^2 = q^2,\; p_n^2 = p^2 \tag{3.18}$$

and the forecasting procedure is in steady state. Then Harrison [5] gives the following algorithm to minimize $V_{1,n} = V_1 = E(e_n^2) = E(y_n - \hat{y}_{n-1}(1))^2$

$$\hat{y}_n(\ell) = \hat{x}_n + \ell\hat{\beta}_n \tag{3.19}$$

$$\hat{x}_n = \hat{x}_{n-1} + \hat{\beta}_{n-1} + G \cdot e_n \tag{3.20}$$

$$\hat{\beta}_n = \hat{\beta}_{n-1} + H \cdot e_n \tag{3.21}$$

where G, H are optimal weights found by solving

$$r^2 = (1-G)\, V_1 \tag{3.22}$$

$$q^2 = (G^2 + GH - 2H)\, V_1 \tag{3.23}$$

$$p^2 = H^2 V_1 \tag{3.24}$$

In Appendix A, we show that (3.22) - (3.24) is equivalent to the steady state solution of (3.7) - (3.12). G,H so obtained are quite different from the Brown [2] algorithms $G = (1-\gamma^2)$, $H = (1-\gamma)^2$ where γ is an arbitrary value.

A heuristic, sub-optimal algorithm was tested in Orr [6] by assuming $p^2 \approx q^2/100$, $r^2/q^2 = k$. This expressed the feeling that the deviation in growth is no more than 10% of deviation in process level. Then from (3.22), (3.24)

$$r^2/p^2 = (1-G)/H^2 = 100k \tag{3.25}$$

from which

$$H^2 = (1-G)/100k \qquad (3.26)$$

G was then updated using (2.5) under the approximation $H \approx o$ and then an estimate of H_n was $\sqrt{(1-G_n)/100k}$.

<u>Moving Average Algorithm</u> - given (3.18) & p = o.

The linear growth model is assumed to have a constant growth β

Algorithm is given by (2.22)

We show in Appendix B that

$$V_{1,\beta} = V_{1,o} + \frac{\beta^2}{4}(B+1)^2 \qquad (3.27)$$

$$V_{L,\beta} = V_{L,o} + L^2 \frac{\beta^2}{4}(B+1)^2$$

$$+ \frac{\beta^2}{4}[L^2(L-1)^2 + 2(B+1)(L-1)(L)(L+1)] \qquad (3.28)$$

where

$$V_{1,\cdot} = E(y_{n+1} - \frac{1}{B}\sum_{j=1}^{B} y_{n-j+1})^2 \qquad (3.29)$$

$$V_{L,\cdot} = E[\sum_{\ell=1}^{L}(y_{n+\ell} - \frac{1}{B}\sum_{j=1}^{B} y_{n-j+1})]^2 \qquad 3.30)$$

$V_{1,o}$, $V_{L,o}$ given by (2.23) and (2.24)

$V_{1,\beta}$, $V_{L,\beta}$ are MSE for linear growth model with constant growth β

We see immediately how much worse MA algorithms do on linear growth models compared to their performance on dynamic mean models - specifically the last terms in (3.27) and (3.28). Expressions for optimal base B^* are quite complex, but one can see that the base period should be shortened as $|\beta|$,L increase.

Note also when p = o, from (3.22)-(3.24), that H = o and $V_1(\text{opt}) = \frac{r^2}{1-G_\infty}$ using (3.19), (3.20) algorithm. This value is lower than that for

a MA algorithm for any β, as expected, since (3.20) is the optimal steady state algorithm.

3.2 Dynamic Proportion Model

This model is useful when expressing noise as a "percentage" of the process level variables.

$$z_n = u_n \cdot \rho_n, \; \rho_n \geq 0$$
$$u_n = u_{n-1} \cdot w_n, \; w_n \geq 0 \tag{3.31}$$

$$E(\rho_n) = E(w_n) = 1 \tag{3.32}$$

where

z_n = observed value of process in period n

u_n = mean of process in period n

ρ_n = multiplicative noise random variable

w_n = multiplicative random change or "percentage" change in in mean u from n-1 to n.

Unlike the Dynamic Mean, here we do not have the possibility of "going negative" on the process variables.

Assume all variables are distributed log normally. Then

$$E(\rho_n) = e^{\beta_\rho + 1/2\, r^2} \tag{3.33}$$

$$E(w_n) = e^{\beta_w + 1/2\, q^2} \tag{3.34}$$

$$Var(\rho_n) = e^{2\beta_\rho + r^2}\,(e^{r^2} - 1) \tag{3.35}$$

$$Var(w_n) = e^{2\beta_w + q^2}\,(e^{q^2} - 1) \tag{3.36}$$

where (β_ρ, r^2), (β_w, q^2) are mean and variance of normal variates $\log \rho$, $\log w$ respectively.

Constraint (3.32) applied to (3.33) - (3.36) yields

$$\beta_\rho = 1/2\ r^2 \tag{3.37}$$

$$\beta_w = 1/2\ q^2 \tag{3.38}$$

$$\mathrm{Var}\ \rho_n = e^{r^2} - 1 \tag{3.39}$$

$$\mathrm{Var}\ w_n = e^{q^2} - 1 \tag{3.40}$$

Letting $y_n = \log_e z_n$ in (3.31) we obtain

$$y_n = x_n' + \beta_\rho + \eta_n,\ E(\eta_n) = o$$

$$x_n' = x_{n-1}' + \beta_w + v_n,\ E(v_n) = o$$

and with $x_n \equiv x_n' + \beta_\rho$,

$$y_n = x_n + \eta_n \tag{3.41}$$

$$x_n = x_{n-1} + \beta_w + v_n \tag{3.42}$$

To obtain the structure of (3.1) we add the identity (3.43)

$$\beta_w = \beta_w + \delta_n \quad \text{with } \mathrm{Var}(\delta_n) = p^2 = o \tag{3.43}$$

$$E(x_o) = m + \beta_\rho \quad \text{where } m = \log_e u_o \tag{3.44}$$

System (3.41) - (3.44) is a special case of the linear growth model with β = constant = $-1/2\ q^2$; previous algorithms are then suitable for forecasting the transformed time series $y_n = \log z_n$ from which $\hat{z}_n = e^{\hat{y}_n}$

If r^2, q^2 are small, we can relax the assumption of log-normality, since the initial relations approximately hold. For example:

$$\log \rho \approx (\rho-1) - 1/2\ (\rho-1)^2 \text{ for } o < \rho \leq 2$$

$$\text{if } E(\rho) = 1,\ E(\log\ ) \approx o - 1/2\ r^2 \approx o$$

$$\mathrm{Var}(\log\rho) \approx \mathrm{Var}(\rho-1) = \mathrm{Var}\,\rho \text{ which is satisfied since } r^2 \approx e^{r^2} - 1$$

PART II

ANCILLARY VARIABLES, TRANSFORMATIONS

CHAPTER IV

PRIMARY ANCILLARIES, COMPARATIVE ANALYSIS OF TRANSFORMATIONS

In this chapter we shall consider models of an observed process D_n (e.g. demand for a repair part), where D_n is related to an independent variable h_n (e.g. hours flown by aircraft using the part). Relations among the parameters of the processes, D_n, D_n/h_n, log D_n and log D_n/h_n will be explored.

4.1 D Versus D/h

If one assumes that the noise terms in equation (2.1) are proportional to the process levels x (which is most reasonable), one obtains quite powerful and flexible versions of the basic model. Also with the incorporation of the ancillary variables h_n into the model, the performance of two Dynamic Mean algorithms can be compared.

$$D_n = x_n \cdot (1+\eta_n) \quad \text{(a)}$$

$$x_n = a_n h_n \quad \text{(b)} \qquad (4.1)$$

$$a_n = a_{n-1}(1+\varepsilon_n) \quad \text{(c)}$$

$$h_n = h_{n-1}(1+\delta_n) \quad \text{(d)}$$

where

D_n = basic observable

x_n = mean of process

a_n = fluctuating rate of x_n/h_n

h_n = ancillary variable

η_n, ε_n, δ_n noise terms with $E(\cdot) = o$, $\mathrm{Var}(\cdot) = V$,

From (4.1)(b), (c), (d),

$$x_n = x_{n-1} \cdot (1 + \varepsilon_n + \delta_n + \delta_n \varepsilon_n) \qquad (4.1)\text{e}$$

Note (4.1)(a) and (4.1)(e) is a form of dynamic mean model. A similar pair of equations can be obtained for the observable D_n/h_n, i.e.,

$$\left.\begin{aligned} D_n/h_n &= a_n(1+\eta_n) \\ a_n &= a_{n-1}(1+\varepsilon_n) \end{aligned}\right\} \qquad (4.2)$$

Now we determine the k factors for D_n and D_n/h_n. First note that

$$\begin{aligned} \text{Var}\ (C_\tau S_t) &= \text{E Var}\ (C_\tau S_t \mid C_\tau) + \text{Var E}\ (C_\tau S_t \mid C_\tau) \\ &= \text{E}\ (C_\tau^2)\ V_s + o \qquad (4,3) \end{aligned}$$

where S_t is a noise variable η, ε, δ and C_τ is any of the multiplier variables in (4.1), (4.2). Hence

$$\begin{aligned} k_{D/h} &= \frac{\text{Var}\ (a_n\eta_n)}{\text{Var}\ (a_{n-1}\varepsilon_n)} = \frac{E(a_n^2)V_\eta}{E(a_{n-1}^2)V_\varepsilon} = \frac{V_\eta}{V_\varepsilon} \cdot \frac{E(a_{n-1}^2)(1+V_\varepsilon)}{E(a_{n-1}^2)} \\ &= \frac{V_\eta}{V_\varepsilon} + V_\eta \qquad (4.4) \end{aligned}$$

$$\begin{aligned} k_D &= \frac{\text{Var}\ (x_n \cdot \eta_n)}{\text{Var}\ (x_{n-1} \cdot (\varepsilon_n + \delta_n + \delta_n\varepsilon_n))} \\ &= \frac{E(x_n^2)}{E(x_{n-1}^2)} \quad \frac{V_\eta}{\underbrace{V_\varepsilon + V_\delta + V_\delta V_\varepsilon}_{V_\nu}} = \frac{E(X_{n-1}^2)(1+V_\nu)}{E(x_{n-1}^2)} \cdot \frac{V_\eta}{V_\nu} \\ &= \frac{V_\eta}{V_\nu} + V_\eta \qquad (4.5) \end{aligned}$$

Typically V_η is quite small (since η_n is a percentage change from mean), so only the first terms of (4.4) and (4.5) are significant.

$$\frac{k_{D/h}}{k_D} \doteq \frac{V_\varepsilon + V_\delta + V_\delta V_\varepsilon}{V_\varepsilon} \qquad (4.6)$$

Therefore $k_{D/h}$ is larger than k_D and can be much larger if $V_\varepsilon << V_\delta$. Let us now determine the impact on the long term mean square error under optimal steady state algorithms as given by equation (2.14). For the D process

$$V^*_{1,D} = E(D_n - \hat{D}_n)^2 = \frac{E(x_n^2)V_\eta}{1 - G_D} \tag{4.7}$$

For the D/h process

$$V^*_{1,D/h} = E(D_n/h_n - (\widehat{D_n/h_n})]^2 = \frac{E(a_n^2)V_\eta}{1 - G_{D/h}} \tag{4.8}$$

The L.H.S. of (4.7) and (4.8) differ by a factor $E(h_n^2)$ (or $E(1/h_n^2)$, depending on the viewpoint) for comparison of forecast errors. Hence for comparative purposes, we compare $V^*_{1,D}$ to $E(h_n^2)\cdot V^*_{1,D/h}$

$$\frac{FE^{[D/h]}}{FE^{[D]}} = \frac{1 - G_D}{1 - G_{D/h}} \tag{4.9}$$

where

$FE^{[P]}$ = squared forecast error of basic observable D using the optimal algorithm on process P

Since $k_{D/h} > k_D$, (2.12) gives $G_{D/h} < G_D$ and hence $FE^{[D/h]} < FE^{[D]}$.

In particular as $k \to \infty$, $G \to k^{-1/2}$. If $k_D = 16$ and $k_{D/h} = 81$, then

$$\frac{1 - 1/4}{1 - 1/9} = \frac{27}{32} \text{ or about 15\% improvement,}$$

using the D/h variable (exact ratio is $\frac{1-.22069}{1-.10511}$)

Result using actual demand data: From Orr [8] mean absolute errors averaged over a group of items gives a ratio which is then squared

$$\frac{(19.834)^2}{10.942)^2} = .808$$

For this group of items $k_D = 4.251$, $k_{D/H} = 14.18$ from which

$$\frac{1 - G_D}{1 - G_{D/h}} = \frac{1 - .38145}{1 - .23263} = \frac{.61854}{.76737} = .806$$

This section and succeeding sections of Chapter IV demonstrate mathematically the intuitive notion that utilizing the ancillary variable, if it is in fact inbedded in the model which properly describes the process, should improve forecasting performance. Also the mathematical groundwork is laid for Chapter V, where only a secondary ancillary is available for observation.

4.2 D vs log D

As in section 4.1,

$$\begin{aligned} D_n &= x_n \quad (1 + \eta_n) \\ x_n &= x_{n-1} \cdot (1 + v_n) \end{aligned} \tag{4.10}$$

from which, as in (4.5)

$$k_D = \frac{V_\eta \; (1+V_v)}{V_v} \tag{4.11}$$

Comparing (4.10) with (3.31)

$$1 + \eta_n = \rho_n, \; (1 + v_n) = w_n \text{ and therefore}$$

$$\begin{aligned} V_\eta &= \operatorname{Var} \rho_n = e^{r^2} - 1 \\ V_v &= \operatorname{Var} w_n = e^{q^2} - 1, \end{aligned} \tag{4.12}$$

From (3.39) and (3.40) equation (4.11) becomes

$$k_D = \frac{e^{r^2} - 1}{e^{q^2} - 1} \cdot e^{q^2} \tag{4.13}$$

From section (3.2) we know

$$k_{\log D} = r^2/q^2 \qquad (4.14)$$

r^2 and q^2 are typically small, and

$$k_D \doteq \frac{r^2}{q^2}\ (1+q^2), \text{using } e^{S^2} \sim 1 + S^2 \qquad (4.15)$$

A comparison of $FE^{[D]}$ vs $FE^{[\log D]}$ cannot be made directly since the optimal algorithms using k_D, $k_{\log D}$ are minimizing different error measures, i.e., $G_{\log D}$ is minimizing $E(\log D - \log F)^2$, which is related to relative measures such as $E(\frac{D-F}{D})^2$, where F is the forecasted $\hat{D}$ obtained by an inverse transformation of the "best" forecast of log D.

4.3 log D versus log D/h

For purpose of this comparison, let's assume the processes can be represented with dynamic proportion models

$$D_n = u_n \rho_n \qquad (a)$$

$$u_n = a_n h_n \qquad (b) \qquad (4.16)$$

$$a_n = a_{n-1} \zeta_n \qquad (c)$$

$$h_n = h_{n-1} \xi_n \qquad (d)$$

where ρ_n, ζ_n, ξ_n are noise variables with expected value = 1

$$u_n = u_{n-1} w_n \text{ where } w_n = \zeta_n \xi_n \qquad (e)$$

Now

$$D_n/h_n = a_n \rho_n \qquad (4.17)$$

Making the log transformation

$$\log D_n = \log u_n + \log \rho_n$$

$$\log u_n = \log u_{n-1} + \log \gamma_n + \log \xi_n \tag{4.18}$$

$$\log D_n/h_n = \log a_n + \log \rho_n$$

$$\log a_n = \log a_{n-1} + \log \gamma_n \tag{4.19}$$

Letting $\upsilon_\cdot = \mathrm{Var}\,(\log(\cdot))$ one finds

$$k_{\log D} = \frac{\upsilon_\rho}{\upsilon_\gamma + \upsilon_\xi} \tag{4.20}$$

$$k_{\log D_h} = \frac{\upsilon_\rho}{\upsilon_\gamma} \tag{4.21}$$

In this case we can make comparisons since

$$V^*_{1,\log D} = E\,(\log D - \log F)^2 = E\,(\log\,(D/F))^2 \tag{4.22}$$

$$V^*_{1,\log D/h} = E\,(\log D/h - \log F/h)^2 = E\,(\log\,(D/F))^2 \tag{4.23}$$

where an inverse transformation is made on the optimal forecast $\log F/h$ to obtain the forecast F of D. Since $k_{\log D/h} > k_{\log D}$, analogously to analysis in section (4.1), we find $V^*_{1,\log D/h}$ to be smaller, i.e.

$$\frac{V^*_{1,\log D/h}}{V^*_{1,\log D}} = \frac{1 - G_{\log D}}{1 - G_{\log D/h}} \tag{4.24}$$

where $\quad G_\cdot = ((1 + 4k_\cdot)^{1/2} - 1)/2k_\cdot \qquad (4.25)$

In Orr [8], it was found that for items with higher activity (larger D), forecasting from log D transformation did better. Arguing from equation (2.14), this could be explained by a modification of the modelling in this section, where the mean rate of log D/h is quite stable for these active items, but that the total inherent noise in the process is then encompassed by a higher variance r^2; this would make $r^2/1\text{-}G$ relatively larger than for the log D process, where r^2 could be assumed smaller with the other variation being explained by fluctuations in the mean of log D.

4.4 D/h versus log D/h

For D/h: The assumed model is

$$D_n h_n = a_n(1 + \eta_n)$$
$$a_n = a_{n-1}(1+\varepsilon_n) \tag{4.26}$$

For log D/h: The assumed model is

$$D_n / h_n = a_n \rho_n$$
$$a_n = a_{n-1} \zeta_n \tag{4.27}$$

We have shown for D/h:

$$r_n^2 = \mathrm{Var}\,(a_n \eta_n) = E(a_n^2)\, V_\eta$$

$$q_n^2 = \mathrm{Var}\,(a_{n-1} \varepsilon_n) = E(a_{n-1}^2)\, V_\varepsilon$$

and for log D/h, from (3.39), (3.40)

$$\mathrm{Var}\ \rho_n = e^{\lambda_n^2} - 1$$

$$\mathrm{Var}\ \zeta_n = e^{\delta_n^2} - 1$$

Also from (4.12) $\mathrm{Var}\,\rho_n = V_\eta$, $\mathrm{Var}\,\zeta_n = V_\varepsilon$

Similarly to section 4.2

$$k_{D/h} = \frac{r_n^2}{q_n^2} = \frac{V_\eta(1+V_\varepsilon)}{V_\varepsilon} \doteq \frac{\lambda_n^2}{\sigma_n^2}\,(1+\sigma_n^2) \tag{4.28}$$

$$k_{\log D/h} = \frac{\lambda_n^2}{\sigma_n^2} \tag{4.29}$$

A useful result is given for use in equation (2.5)

$$G_{n+1} = \frac{q_n^2 + r_n^2\,G_n}{q_n^2 + r_n^2\,G_n + r_{n+1}^2}$$

If $r_n^2 \propto 1/h_n^2$ and $E(a_n^2)$ is independent of h_n (both reasonable) then

$$e^{\lambda_n^2} - 1 = \mathrm{Var}\,\rho_n = V_\eta \propto 1/h_n^2$$

and for λ_n^2 small, $\lambda_n^2 \propto 1/h_n^2$. Therefore $r_n^2 h_n^2$ and $\lambda_n^2 h_n^2$ are constant and equation (2.5) can be written in terms of $k_{D/h}$ or $k_{\log D/h}$ as

$$G_{n+1} = \frac{1 + k\,G_n}{1 + k\,G_n + k\cdot h_n^2/h_{n+1}^2} \tag{4.30}$$

Therefore the weight G_n applied to current observations when forecasting D/h or log D/h increases in some relation to current value squared of the ancillary variable h_n.

CHAPTER V

SECONDARY ANCILLARIES

5.1 Focus

In this chapter, we discuss the use of secondary ancillaries - independent variables which are less directly related than the primaries to the basic observable being forecasted. In many cases one must utilize observations of these secondary variables when the primary data is not available. For example D_n, the basic variable being forecast, is demand for a repair part; h_n, the primary variable, is a measure of usage such as total hours flown by airplanes using the part; and the secondary ancillary, designated g_n, might be the number of these airplanes ("density") available for flight during the period n. Usage data may not be measured or retained on file, whereas the density information is readily available.

We have found, in general, it is desirable to apply optimal algorithms to processes which have high k-factors. If there exists a relation between demand D and usage h, then $k_{D/h} > k_D$. If, in turn, there is a dependency of usage on density g, in that a certain percentage of the variance in h can be explained by variance in g, then variation in D is due in part to density g. However the direct fluctuation of D with h has been obscured and more noise is present, and hence D/g is a less stationary or stable "rate" variable than is D/h. We would expect to find these inequalities,

$$k_{D/h} > k_{D/g} > k_D$$

We know from Chapter IV that forecasting from a process with relatively higher k yields better performance in projecting the ultimate variable D. Future work can be done to compare the relative performances for D/h, D/g, D when:

i) g is observable over all past periods and h is only intermittently observable.

ii) The models of the ancillary processes g and h are such that their forecasts for future periods are limited in accuracy and hence

$$\widehat{(D/h)} \times \text{forecasted } h \quad \text{and}$$

$$\widehat{(D/g)} \times \text{forecasted } g$$

are degraded forecasts of the variable D.

5.2 Model & Comparison of K-factors

Let us rewrite equations (4.1) and (4.2) and append some similar modeling on a secondary variable g.

$$\begin{aligned}
D_n &= x_n\,(1+\gamma_n) && \text{(a)}\\
x_n &= a_n h_n && \text{(b)}\\
a_n &= a_{n-1}(1+\varepsilon_n) && \text{(c)} \qquad (5.1)\\
h_n &= h_{n-1}\,(1+\delta_n) && \text{(d)}\\
h_n &= b_n g_n && \text{(e)}\\
g_n &= g_{n-1}(1+w_n) && \text{(f)}\\
b_n &= b_{n-1}(1+\theta_n) && \text{(g)}
\end{aligned}$$

where b_n = fluctuating rate coefficient for usage per density.

All noise terms, $\gamma \ldots \theta$, have $E(\cdot) = o$, $Var(\cdot) \equiv V$. The equations below follow immediately from the above.

$$\begin{aligned}
x_n &= x_{n-1}\,(1 + \varepsilon_n + \delta_n + \delta_n \varepsilon_n) && \text{(h)}\\
D_n/h_n &= a_n(1+\gamma_n) && \text{(i)}\\
h_n &= h_{n-1}\,(1 + \theta_n + w_n + w_n \theta_n) && \text{(j)} \qquad (5.1)\\
D_n/g_n &= a_n b_n\,(1+\gamma_n) && \text{(k)}\\
a_n b_n &= a_{n-1} b_{n-1}(1 + \varepsilon_n + \theta_n + \varepsilon_n \theta_n) && \text{(l)}
\end{aligned}$$

For reference we rewrite (4.4) and (4.5)

$$k_{D/h} = \frac{V_\eta}{V_\varepsilon} + V_\eta \doteq \frac{V_\eta}{V_\varepsilon} \tag{5.2}$$

$$k_D = \frac{V_\eta}{V_\varepsilon + V_\zeta + V_\zeta V_\varepsilon} + V_\eta \doteq \frac{V_\eta}{V_\varepsilon + V_\zeta + V_\zeta V_\varepsilon} \tag{5.3}$$

Similarly,

$$k_{D/g} = \frac{E(a_n b_n)^2 V_\eta}{E(a_{n-1} b_{n-1})^2 (V_\varepsilon + V_\theta + V_\varepsilon V_\theta)} \doteq \frac{V_\eta}{V_\varepsilon + V_\theta + V_\varepsilon V_\theta} \tag{5.4}$$

From (5.2), (5.3), (5.4) approximations,

$$1/k_D = 1/k_{D/h} + \frac{V_\zeta + V_\zeta V_\varepsilon}{V_\eta} \tag{5.5}$$

$$1/k_{D/g} = 1/k_{D/h} + \frac{V_\theta + V_\theta V_\varepsilon}{V_\eta} \tag{5.6}$$

Equations (5.5) and (5.6) give

$$\frac{1/k_D - 1/k_{D/h}}{1/k_{D/g} - 1/k_{D/h}} = \frac{V_\zeta}{V_\theta} \frac{(1+V_\varepsilon)}{(1+V_\varepsilon)} = \frac{V_\theta + V_\omega + V_\omega V_\theta}{V_\theta} \tag{5.7}$$

R.H.S. of (5.7) was obtained from (5.1)(d) and (j) whence $\zeta = \theta + \omega + \omega\theta$

We now derive a formula from (5.7) for obtaining $k_{D/g}$ when k_D and $k_{D/h}$ are known.

Equation (5.7) can be rewritten

$$\frac{k_{D/g}(k_{D/h} - k_D)}{k_D(k_{D/h} - k_{D/g})} = 1 + \frac{V_\omega}{V_\theta}(1 + V_\theta) = 1 + F \tag{5.8}$$

Let $k_{D/g} = \phi k_D$ and solve (5.8) for ϕ

$$\phi = \frac{(1+F)\, k_{D/h}}{k_{D/h} + F\, k_D} \tag{5.9}$$

Note if $V_\omega = o$, $F = o$ and $\phi = 1 \Rightarrow k_{D/g} = k_D$. This shows that if density g_n does not vary, it is useless as an aid to forecasting and one may just as well operate on D_n as D_n/g_n.

Note if $V_\theta = 0$, $F = \infty$, and $\phi = \frac{k_{D/h}}{k_D} \Rightarrow k_{D/g} = k_{D/h}$

That is, if b_n is a constant then g_n is as good an auxiliary variable as h_n and D_n/g_n forecasts as well as D_n/h_n.

Equation (5.9) was utilized in conjunction with empirical estimates of V_ω, V_θ to obtain the $k_{D/g}$ factors in Orr [8].

5.3 Estimates and Empirical Results

From (5.1)(f)

$$\text{Var}\,(g_{n-1} w_n) = \text{Var}\,(g_n - g_{n-1})$$

and using equation (4.3),

$$V_w \equiv \text{Var}\, w_n = \frac{\text{Var}\,(g_n - g_{n-1})}{E(g_{n-1}^2)} \tag{5.10}$$

An estimator for (5.10) using a series of N observations on g

$$\hat{V}_w = \frac{\frac{1}{N-2} \sum^{N-1} (g_i - g_{i-1})^2}{\frac{1}{N} \sum^{N} (g_i^2)} \tag{5.11}$$

In Orr [8] there were 20 time series on aircraft density for the 10,000 repair parts investigated, and $\hat{V}_w$ was averaged over 20 estimates to give $\hat{\hat{V}}_w$

From (5.1)(e) and (g)

$$\mathrm{Var}\,(b_{n-1}\theta_n) = \mathrm{Var}\,(b_n - b_{n-1})$$

$$V_\theta = \frac{\mathrm{Var}\,(h_n/g_n - h_{n-1}/g_{n-1})}{E((h_{n-1}/g_{n-1})^2)} \tag{5.12}$$

$$\hat{V}_\theta \approx \frac{1/N\text{-}2 \sum^{N-1} (h_i/g_i - h_{i-1}/g_{i-1})^2}{1/N \sum^{N} (h_i/g_i)^2} \tag{5.13}$$

Again in Orr [8], a refined estimate $\hat{\hat{V}}_\theta$ was used.

Empirical Results: Data used by Orr [8] gives

$$\hat{\hat{V}}_w \doteq .01526 \qquad \hat{\hat{V}}_\theta \doteq .02613$$

From (5.8), $F \doteq .6$ and hence

$$\phi = \frac{1.6\, k_{D/h}}{k_{D/h} + .6\, k_D}$$

$$k_{D/g} = \phi\, k_D$$

$k_{D/g}$ is at most 1.6 times larger than k_D even though $k_{D/h}$ on this data could become quite large. The coefficient of determination R^2 equaled .72 in a regression of usage h versus density g on this data. A relation can be derived between R^2 and V_w, V_θ but is not presented here; the value of .72 is consistent with the estimates of V_w, V_θ.

BIBLIOGRAPHY

1. Box & Jenkins, Time Series Analysis: Forecasting & Control, Holden-Day 1970.

2. Brown, R., Smoothing, Forecasting, & Prediction, Prentice-Hall 1963.

3. Cogger, K.D. "The Optimality of General-Order Exponential Smoothing," Operations Research, V. 44 #4, July-Aug 1974.

4. Goodman, M.L., "A New Look at Higher Order Exponential Smoothing for Forecasting," Operations Research, V. 44 #4, July-Aug 1974.

5. Harrison, P.J., "Exponential Smoothing & Short Term Sales Forecasting," Management Science, July 1967.

6. Jewell, W.J., "The Credible Distribution," OR Center, Report #77-13, University of California, Berkeley, Aug 1973.

7. Kalman, R.E. and R.S. Bucy, "New Results in Linear Filtering and Prediction Theory," J. Basic Engrg, V.83, Dec 1961, p.662-680.

8. Orr, D.A., "Demand Forecasts Using Process Models & Item Class Parameters: Application of Ancillary Variables," USDRC Inventory Research Office, IRO 219, April 1976, AD#A026081.

9. Sage, A.P., Optimum Systems Control, Prentice Hall, 1968.

APPENDIX A

EQUIVALENT ALGORITHMS

In this appendix we show that seemingly different forecast algorithms presented in the literature are equivalent in general or in special cases. One would only expect that if the underlying models of the process are the same and/or generate the same observables in a statistical sense and if the minimizing criterion (e.g. mean square error of one-period forecast) is the same, the optimal algorithms can be shown to be identical. In the following sections we shall do some pairwise comparisons of algorithms developed or presented by Kalman [7], Harrison [5], Box and Jenkins [1], Jewell [6] and of algorithms obtained by regression and spectral techniques.

A.1 Kalman Versus Harrison

We have shown in the main report, Section 2.1, that the Kalman G_n approaches the G_∞ in (2.12) given by Harrison for the dynamic mean model with constant r^2, q^2. In steady-state operation, constant weighting factor G_∞ minimizes MSE $V_1 = \sigma_\infty^2 + q^2 + r^2$ (from 2.21) where σ_∞^2 = Var (x_∞|all past history).

For the linear growth model of Section (3.1), we now show the Kalman 2nd order weights G_n, H_n approach G,H which are solutions of Harrison's equations (3.22), (3.23), (3.24).

In equations (3.7) - (3.12) drop the subscript n and find steady-state relations

$$V = s_{11} + r^2 \tag{A1}$$

$$G = s_{11}/V \tag{A2}$$

$$H = s_{12}/V \tag{A3}$$

$$0 = - G^2V + 2s_{12} - 2\,GHV + s_{22} - H^2V + q^2 + p^2 \tag{A4}$$

$$0 = - GHV + s_{22} - H^2V + p^2 \tag{A5}$$

$$0 = - H^2V + p^2 \tag{A6}$$

From (A6), (3.24) follows. (A6) and (A5) give $s_{22} = GHV$ and (A4) now becomes

$$0 = -G^2V + 2s_{12} - GHV + q^2 \qquad \text{(A7)}$$

(A3) used in (A7) yields

$$0 = -G^2V + 2HV = GHV + q^2 \qquad \text{(A8)}$$

from which (3.23) follows. Inserting $s_{11} = GV$ from (A2) in (A1) easily gives (3.22).

A.2 Harrison Versus Box-Jenkins

In this section we show that the dynamic mean and linear growth processes can be reformulated as integrated moving average (IMA) processes, (0,1,1) and (0,2,2) respectively (Box-Jenkins notation). The steady-state optimal algorithms in sections 2.1 and 3.1 are equivalent to the forecast procedures in B-J [1 , pp. 144, 147].

Dynamic Mean vs IMA (0,1,1):

Equation 2.1 can be expressed by

$$y_n = y_{n-1} + \gamma_n - \gamma_{n-1} + v_n \qquad \text{(A9)}$$

$$w_n \equiv y_n - y_{n-1} = \gamma_n - \gamma_{n-1} + v_n \qquad \text{(A10)}$$

Expressed as an IMA, (A10) is

$$w_n \equiv y_n - y_{n-1} = a_n - \theta a_{n-1}, \qquad \text{(A11)}$$

where a = random stock with $E(a) = o$, $E(a^2) = \sigma_a^2$

B-J forecasting from (A11) is given by

$$\hat{y}_{n-1}(1) = y_{n-1} - \theta a_{n-1} \qquad \text{A12)}$$

Box-Jenkins shows that under optimal forecasting, minimizing MSE, that the residuals a_n are the one-step ahead forecast errors

$$y_n - \hat{y}_{n-1}(1) = a_n \tag{A13}$$

Combining (A12), (A13)

$$\hat{y}_{n-1}(1) = y_{n-1} - \theta\,(y_{n-1} - \hat{y}_{n-2}(1))$$

$$= \hat{y}_{n-2}(1) + (1-\theta)\,[y_{n-1} - \hat{y}_{n-2}(1)] \tag{A14}$$

Note (A14) is of same form as (2.11) with $1-\theta = G_\infty$

Now from (A10), A11)

$$E(w_i^2) = r^2 + r^2 + q^2 = (1+\theta^2)\,\sigma_a^2 \tag{A15}$$

$$E(w_i w_{i-1}) = -r^2 = -\theta\sigma_a^2 \tag{A16}$$

Therefore autocorrelation of lag 1,

$$P_1 = \frac{-r^2}{2r^2+q^2} = \frac{-\theta}{1+\theta^2} \tag{A17}$$

Solving (A17) for θ in terms of r^2, q^2

$$\theta_\pm = \frac{2r^2+q^2}{2r^2} \pm \frac{1}{2}\sqrt{\left(\frac{(2r^2+q^2)}{r^2}\right)^2 - 4}$$

θ is required to be < 1. We ignore θ_+, being > 1.

$$\theta_- = \frac{2r^2+q^2}{2r^2} - \sqrt{\frac{4\,r^2q^2 + q^4}{4r^2}}$$

$$= \frac{2k+1}{2k} - \frac{1}{2k}\sqrt{4k+1} = \frac{2k+1-\sqrt{4k+1}}{2k} \tag{A18}$$

and $1 - \theta_- = (\sqrt{4k+1} - 1)/2k$ QED

Linear Growth vs IMA (0,2,2):

Equation 3.1 can be expressed by

$$w_n \equiv y_n - y_{n-1} = \eta_{n-1} + \beta_n + v_n + \eta_n$$

$$Z_n \equiv w_n - w_{n-1} = \eta_n - \eta_{n-1} - \eta_{n-1} + \eta_{n-2} + v_n - v_{n-1} + \delta_n \tag{A18}$$

Expressed as an IMA, (A19) is

$$y_n - 2y_{n-1} + y_{n+2} \equiv Z_n = a_n - \theta_1 a_{n-1} - \theta_2 a_{n-2} \tag{A20}$$

Box-Jenkins gives an integrated form of their forecast which is structurally equivalent to (3.19) - (3.21) with

$$G' = 1 + \theta_2 \text{ and } H' = 1 - \theta_1 - \theta_2 \tag{A21}$$

We now show that G', H' are solutions of (3.22)-(3.24). From (A19), (A20),

$$E(Z_i^2) = r^2 + 4r^2 + r^2 + q^2 + q^2 + p^2 = (1+\theta_1^2+\theta_2^2)\sigma_a^2 \tag{A22}$$

$$E(Z_i Z_{i-1}) = E[(\eta_i - 2\eta_{i-1} + \eta_{i-2} + v_i - v_{i-1} + \delta_i)$$

$$\cdot(\eta_{i-1} - 2\eta_{i-2} + \eta_{i-3} + v_{i-1} - v_{i-2} + \delta_{i-1})]$$

$$= E[(a_i - \theta_1 a_{i-1} - \theta_2 a_{i-2})(a_{i-1} - \theta_1 a_{i-2} - \theta_2 a_{i-3})]$$

$$= -2r^2 - 2r^2 - q^2 = (-\theta_1 + \theta_1\theta_2)\sigma_a^2 \tag{A23}$$

Similarly

$$E(Z_i Z_{i-2}) = r^2 = -\theta_2\sigma_a^2 \tag{A24}$$

From (A23), (A24) and (A22), autocorrelations of lag 1,2

$$\rho_1 = \frac{-4r^2 - q^2}{6r^2 + 2q^2 + p^2} = \frac{-\theta_1(1-\theta_2)}{1 + \theta_1^2 + \theta_2^2} \tag{A25}$$

$$\rho_2 = \frac{r^2}{6r^2 + 2q^2 + p^2} = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} \tag{A26}$$

For optimal forecasts, the MSE of 1-step forecasts

$$E(y_n - \hat{y}_{n-1}(1))^2 = V = \sigma_a^2 \tag{A27}$$

Therefore from (A22)

$$6r^2 + 2q^2 + p^2 = (1 + \theta_1^2 + \theta_2^2)\, V \tag{A28}$$

Using (A28) in (A25), (26)

$$-4r^2 - q^2 = -\,\theta_1\,(1 - \theta_2)\, V \tag{A29}$$

$$r^2 = -\,\theta_2\, V \tag{A30}$$

(3.22) follows from (A30) and (A21).

Solving (A29), (A30) for q^2

$$q^2 = (4\,\theta_2 + \theta_1 - \theta_1\theta_2)\, V \tag{A31}$$

Noting from (A21) $-\theta_2 = 1 - G'$, $\theta_1 = 2 - G' - H'$,

$$q^2 = (-4(1-G') + 2 - G' - H' + (1-G')(2-G'-H'))$$

$$= (-\,2H' + G'^2 + G'H') \tag{A32}$$

which agrees with (3.23). Finally from (A28), (A30), (A31)

$$p^2 = (1 + \theta_1{}^2 + \theta_2{}^2)\ V - 6r^2 - 2q^2$$

$$= (1 + \theta_1{}^2 + \theta_2{}^2 + 6\theta_2 - 2\ (4\theta_2 + \theta_1 - \theta_1\theta_2)]V$$

$$p^2 = [1 + \theta_1{}^2 + \theta_2{}^2 - 2\ (\theta_1 + \theta_2 - \theta_1\theta_2)]\ V$$

$$= (1 - \theta_1 - \theta_2)^2\ V = H'V \qquad \text{(A32)}$$

QED

Harrison [5], Cogger [3], and Goodman [4] have found some generalizations relating to higher order growth processes

$$y_n = x_n{}^{(1)} + \gamma_n$$

$$x_n{}^{(i)} = x_{n-1}{}^{(i)} + x_n{}^{(i+1)} + v_n{}^{(i)}\ ,\ i = 1 \ldots k$$

and the IMA (o,k,k) structure. The equations exist in the above references for finding the relation amongst optimal weighting factors G, H...; variances of noise γ, $v^{(i)}$; and the IMA coefficients θ_i. But explicit useful formulas have not been presented. Harrison's general solutions, for example, are not immediately translatable into formulae involving θ_i.

More importantly, one should compare the power and flexibility of the general vector model and its Kalman filter forecasts,

$$\underset{\sim}{y}_n = H_n \underset{\sim}{x}_n + \underset{\sim}{\gamma}_n \qquad \text{(A33)}$$

$$\underset{\sim}{x}_n = A_{n-1} \underset{\sim}{x}_{n-1} + \Gamma_n \underset{\sim}{v}_n\ ,$$

With the general ARIMA model

$$\Phi(B)\, y_n = \theta(B)\, a_n \qquad \text{(A34)}$$

where $\phi(\cdot)$, $\theta(\cdot)$ are polynomials in the backward shift operator B. It would be interesting to show the equivalence of the forecasts in steady-state operation.

The approach is as follows:

Let $y_n = x_n + b_n$ and then from (A34)

$$\phi(B)x_n = \theta(B)a_n - \phi(B)b_n \tag{A35}$$

$$= \theta'(B)\, v_n \tag{A36}$$

where $\theta'(B)v_n$ has the same statistical properties (auto-covariance) as RHS of (A35). Now (A36) can be put in the form of (A33) where $\underset{\sim}{x}_n = [x_n x_{n-1} \ldots x_{n-p-1}]^T$ and $\phi(B)$ is of degree p.

$$y_n = \overbrace{[1\ 0\ 0\ \ldots 0]}^{p \text{ elements}}\, \underset{\sim}{x}_n + b_n \tag{A37}$$

$$\underset{\sim}{x}_n = A\,(\phi)\, \underset{\sim}{x}_{n-1} + \Gamma(\theta')\, \underset{\sim}{v}_n$$

In steady-state, Kalman filtering yields

$$\hat{\underset{\sim}{x}}_n = A\, \hat{\underset{\sim}{x}}_{n-1} + \underset{\sim}{G}_\infty (y_n - HA\, \hat{\underset{\sim}{x}}_{n-1}) \tag{A38}$$

$\underset{\sim}{G}_\infty$ is a column vector

$$\hat{y}_n(1) = HA\, \hat{\underset{\sim}{x}}_n,\ \hat{y}_n(\ell) = HA^\ell\ \hat{\underset{\sim}{x}}_n \tag{A39}$$

In (A38), if one-step ahead forecast is truly the same as the B-J forecast the last term becomes $\underset{\sim}{G}_\infty \cdot a_n$ from (A13). Moreover Box-Jenkins procedure satisfies

$$\begin{aligned} \hat{y}_n(1) &= \hat{y}_{n-1}(2) + (\phi_1 - \theta_1)\, a_n \\ \hat{y}_n(\ell) &= \hat{y}_{n-1}(\ell+1) + \psi_\ell\, a_n \end{aligned} \tag{A40}$$

where ψ_ℓ is function of ϕ, θ coefficients.

From (A38), (A39), (A40)

$$HA^{\ell} \hat{\underset{\sim}{x}}_n = HA^{\ell+1} \hat{\underset{\sim}{x}}_{n-1} + \psi_{\ell} a_n \tag{A41}$$

L.H.S. of (A41), using (A38) becomes

$HA^{\ell} (A \hat{\underset{\sim}{x}}_{n-1} + \underset{\sim}{G}_{\infty} a_n)$ and we see that for equality in (A41) that

$$HA^{\ell} \underset{\sim}{G}_{\infty} = \psi_{\ell} \tag{A42}$$

Equation (A42) needs to be proved.

A.3 Kalman Versus Jewell's Credible Mean

Consider the process

$$y_n = x_n + \eta_n, \text{ Var } \eta = r^2$$

$$x_n = x_{n-1} = x \tag{A43}$$

The mean of the process does not change, i.e., $k = \infty$. Consider a prior distribution on x, $\mathcal{D}(x_o; \mu, \tau^2)$ with mean μ and variance τ^2. Now Jewell [6] defines a "mean-credible time constant"

$$\mathcal{N} = \frac{E_{\theta} \text{Var } (y|\theta)}{\text{Var}_{\theta} \, E(y|\theta)} \tag{A44}$$

where process y depends on parameter θ which has a prior distribution. In this case then

$$\mathcal{N} = \frac{E_x \text{ Var } (y|x)}{\text{Var}_x \, E(y|x)} = \frac{E_x(r^2)}{\text{Var}_x(x)} = \frac{r^2}{\tau^2} \tag{A45}^*$$

---

* Another interesting case of (A44) is when Var(y) = C E(y) where C = constant VMR. Then

$$= \frac{E_x(CX)}{\text{Var}_x(x)} = \text{VMR}(y) \cdot \frac{E_x(x)}{\text{Var}_x(x)} = \frac{\text{VMR}(y)}{\text{VMR}(x_o)}$$

Jewell [6] shows the best linear forecast, minimizing MSE, is

$$\hat{y}_n(1) = \mu + \left(\frac{nA}{1+nA}\right)(\bar{y} - \mu) \tag{A46}$$

where $A = 1/\nu$, $\bar{y}$ = sample mean $= \frac{1}{n}\sum_{1}^{n} y_i$

Now we show that the same forecast is obtained after n steps in the operation of the Kalman algorithm (2.3)-(2.7). In (2.5) for $q^2 = o$, r^2 constant,

$$G_{i+1} = \frac{G_i}{G_i+1} \tag{A47}$$

Slightly modifying (2.6), (2.7) for consistent notation

$$\hat{x}_1 = \mu + G_1(y_1 - \mu)$$

$$G_1 = \frac{\tau^2}{\tau^2+r^2} = \frac{\tau^2/r^2}{\tau^2/r^2+1} = \frac{G_o}{G_o+1} \tag{A48}$$

Therefore $G_o = A$. We use an inductive proof

$$\hat{x}_1 = \frac{1}{1+G_o}\mu + \frac{G_o}{1+G_o} y_1$$

$$\hat{x}_2 = \frac{1}{1+G_1}\hat{x}_1 + \frac{G_1}{1+G_1} y_2$$

$$= \frac{1}{1+G_o/(1+G_o)}\left(\frac{1}{1+G_o}\mu + \frac{G_o}{1+G_o} y_1\right) + \frac{G_o/1+G_o}{1+G_o/(1+G_o)} y_2$$

$$= \frac{1}{1+2G_o}\mu + \frac{G_o}{1+2G_o} y_1 + \frac{G_o}{1+2G_o} y_2$$

$$= \mu + \frac{2G_o}{1+2G_o}\left(\frac{y_1+y_2}{2} - \mu\right) \tag{A48}$$

(A48) is of the form (A46). Assume form holds for i^{th} step.

$$x_i = \frac{1}{1+iG_o}\mu + \frac{iG_o}{1+iG_o}\left(\frac{1}{i}\ \underset{\sim}{y}^i\right) \tag{A49}$$

where $\underset{\sim}{y}^i = \sum^i y_j$

Note if $G_{i-1} = \frac{G_o}{1+(i-1)G_o}$, then from (A47)

$$G_i = \frac{G_o/1+(i-1)G_o}{1+G_o/1+(i-1)G_o} = \frac{G_o}{1+iG_o}$$

Now,

$$\hat{x}_{i+1} = \left(1 - G_{i+1}\right)\hat{x}_i + G_{i+1}\, y_{i+1} \qquad \text{from (2.4)}$$

$$= \left(1 - \frac{G_o/1+iG_o}{1+G_o/(1+iG_o)}\right)\left(\frac{\mu}{1+iG_o} + \frac{iG_o}{1+iG_o}\ \frac{1}{i}\ \underset{\sim}{y}^i\right)$$

$$+ \frac{G_o/1+iG_o}{1+G_o/1+iG_o}\ (y_{i+1})$$

$$\hat{x}_{i+1} = \frac{1}{1+(i+1)G_o}\mu + \frac{1}{1+G_o/(1+iG_o)}\ \frac{iG_o}{1+iG_o}\ \frac{1}{i}\ \underset{\sim}{y}^i + \frac{G_o}{1+(i+1)G_o}\ y_{i+1}$$

$$= \quad '' \quad + \frac{iG_o}{1+(i+1)G_o}\ \frac{1}{i}\ \underset{\sim}{y}^i \quad + \quad ''$$

$$= \quad '' \quad + \frac{1}{1+(i+1)G_o}\ (G_o\underset{\sim}{y}^i + G_o y_{i+1})$$

$$= \frac{1}{1+(i+1)G_o}\mu + \frac{(i+1)G_o}{1+(i+1)G_o}\left(\frac{1}{i+1}\ y^{i+1}\right) \tag{A50}$$

QED

A.4 Kalman Versus Regression Analysis, Least Squares & Minimum Variance Estimators

Assume a process

$$D_n = a_n h_n + \gamma'_n \tag{A51}$$

$$a_n = a_{n-1} + \varepsilon_n \tag{A52}$$

(A51) is a single variable regression equation. A vector version of the Kalman algorithm can handle the multi-variable case.

Dividing (A51) by h_n

$$D_n/h_n \equiv y_n = a_n + \gamma_n \tag{A53}$$

where $\mathrm{Var}\,\gamma_n = \mathrm{Var}\,\gamma'_n/h_n^2$. Assume a constant ratio

$$\frac{\mathrm{Var}\,\gamma_n}{\mathrm{Var}\,\varepsilon_n} = \frac{r_n^2}{q_n^2} \equiv k \tag{A54}$$

Assume $\mathrm{Var}\,\gamma'_n$ is not dependent on h_n (homoscedasdicity). Then $r_n^2 h_n^2 = r_{n+1}^2 h_{n+1}^2 = \text{constant } C$ (A55)

Equation (2.5) then becomes

$$G_{n+1} = \frac{1+kG_n}{1+kG_n+kh_n^2/h_{n+1}^2} \tag{A56}$$

Now let $k = \infty$, for which the process is now

$$y_n = a + \gamma_n \tag{A57}$$

with $\quad \mathrm{Var}\ \gamma_n = C/h_n^2 = \mathrm{Var}\ y_n$ (A58)

$$G_{n+1} = \frac{G_n}{G_n+h_n^2/h_{n+1}^2} \tag{A59}$$

Letting $G_o = 1$, then Kalman filter estimates iterate thusly:

$$\hat{y}_1 = y_o + \frac{1}{1+h_o^2/h_1^2}(y_1 - y_o) = \frac{h_o^2}{h_o^2+h_1^2}\ y_o + \frac{h_1^2}{h_o^2+h_1^2}\ y_1$$

$$y_2 = y_1 + \frac{h_1^2/(h_o^2+h_1^2)}{h_1^2/(h_o^2+h_1^2) + h_1^2/h_2^2}\ (y_2 - y_1)$$

$$= \frac{h_2^2}{h_o^2+h_1^2+h_2^2}\ y_2 + \frac{h_o^2+h_1^2}{h_o^2+h_1^2+h_2^2}\ y_1$$

$$= \frac{1}{\sum_o^2 h_i^2}\ (h_2^2 y_2 + h_1^2 y_1 + h_o^2 y_o)$$

and in general

$$y_m = \frac{\sum^m h_i^2 y_i}{\sum^m h_i^2} \tag{A60}$$

y_m is an estimate after m observations of a in the regression equation $D_n = ah_n + \gamma_n$. (A60) may be rewritten

$$y_m = \frac{\sum h_i^2\ D_i/h_i}{\sum h_i^2} = \frac{\sum h_i D_i}{\sum h_i^2} = \hat{a} \tag{A61}$$

(A61) is the estimate of the coefficient using regression analysis. This is expected since, as Sage [9] and others have shown, minimizing by least squares with weight $1/r_n^2$ to determine $\hat{a}$ gives the minimum MSE* if the estimate is linear in y, e.g. as in (A60). This is also the minimum variance estimator as is now shown. (A61) is written

$$y_m = \sum^m w_i y_i \tag{A62}$$

*If in addition, γ_n is Gaussian, then $\hat{a}$ is the maximum likelihood estimate also.

with $w_i = h_i^2/\Sigma h_i^2$ and $\Sigma w_i = 1$

To minimize variance of $\hat{y}_m$, form a Lagrangian function

$$\min_{w,\lambda} \left(\Sigma w_i^2 \operatorname{Var} y_i + \lambda(1 - \Sigma w_i) \right) \tag{A63}$$

The conditions for maximum are

$$2 w_i \operatorname{Var} y_i - \lambda = o \quad \text{or} \quad w_i = \frac{\lambda}{2} \frac{1}{\operatorname{Var} y_i}$$

and since $\Sigma w_i = 1$, $\frac{\lambda}{2} = \frac{1}{\Sigma 1/\operatorname{Var} y_i}$. Hence

$$w_i = \frac{1/\operatorname{Var} y_i}{\Sigma 1/\operatorname{Var} y_i} = \frac{h_i^2/C}{\Sigma h_i^2/C} \quad \text{from (A58)}$$

Note also that (A60) is in the form of a weighted moving average on y.

A.5 Kalman Versus Spectral Analysis (Weiner-Hopf)

The Weiner filter problem, solved in the frequency domain, and the Kalman filter problem solved in the time domain (its power is most apparent for non-stationary problems) must obtain identical results for stationary systems with infinite observation time, since MSE is the minimization criterion in both. I can only paraphrase (and therefore will not do so) the note by R.J. Leake, "Duality Condition Established in the Frequency Domain", IEEE Transactions on Automatic Control, 1965. A more obtuse proof is given in Sage [9], Chapter 9.

APPENDIX B

MEAN SQUARE ERRORS OF MOVING AVERAGE FORECASTS ON THE DYNAMIC MEAN & LINEAR GROWTH MODELS

B.1 Dynamic Mean Case

$$y_n = x_n + \eta_n \text{ , Var } \eta_n = r$$

$$x_n = x_{n-1} + v_n, \text{ Var } v_n = q$$

For simplicity assume $x_o = o$. This doesn't affect generality. We note some relations that will be useful.

$$V\ y_n = E(y_n^2) = V\ x_n + r \tag{B1}$$

$$V\ x_n = q + V\ x_{n-1} = E(x_n^2) \tag{B2}$$

$$E(y_n y_{n-k}) = V\ (x_{n-k}) = E\ (x_n y_{n-k}) \tag{B3}$$

$$E(y_n) = E(x_n) = o \tag{B4}$$

$$V\ (y_n + y_{n-1}) = E(y_n + y_{n-1})^2 = q + 2r + 4\ V\ x_{n-1} \tag{B5}$$

Now

$$(y_{B+1} - \frac{1}{B} \sum_1^B y_i) = \sum_1^{B+1} v_j + \eta_{B+1} - \frac{1}{B} (\sum^B \eta_j + \sum_1^{B+1} (B+1-j) v_j) \tag{B6}$$

$$(\quad " \quad) = \eta_{B+1} - \frac{1}{B} \sum^B \eta_j + \frac{1}{B} \sum^{B+1} (j-1)\ v_j \tag{B7}$$

$$V_1 \equiv E(\quad " \quad)^2 = r + \frac{r}{B} + \frac{q}{B^2} \sum_1^B j^2 \tag{B8}$$

$$V_1 = r + r/B + q\,\frac{(B+1)(2B+1)}{6B} \tag{B9}$$

(B9) $\Rightarrow$ Equation (2.23) QED

To find V_L in Equation (2.24), define

$$\xi_\ell = y_\ell - \hat{x}_0$$

$$\xi_\ell = y_\ell - \frac{1}{B}\hat{\sum} y_{-j}\ ,\ \ell = o \ldots L \tag{B10}$$

Then $$\xi_\ell = \xi_o + (y_\ell - y_o) \tag{B11}$$

$$E(\xi_o^2) = V_1 = r + T \qquad \text{(see (B9))} \tag{B12}$$

$$E(\xi_\ell^2) = E(\xi_o^2) + 2\,E[\xi_o(y_\ell - y_o)] + E(y_\ell - y_o)^2 \tag{B13}$$

Second term is $2E[\,(y_o - \hat{x}_o)\,y_\ell - (y_o - \hat{x}_o)y_o]$

$$= 2\,[V\,x_o - E\,(\hat{x}_o y_\ell) - (V\,x_o + r) + E(\hat{x}_o y_o)]$$

$$= -\,2r \quad \text{using (B1) - (B5)} \tag{B14}$$

Third term is $V\,x_\ell + r + V\,x_o + r - 2\,V\,x_o$

$$= \ell q + 2r \tag{B15}$$

Therefore $$E(\xi_\ell^2) = E(\xi_o^2) + q \tag{B16}$$

Similarly for $k < \ell$, $k \neq o$

$$E(\xi_\ell \xi_k) = E[(\xi_o + y_\ell - y_o)\,(\xi_o + y_k - y_o)]$$

$$= E(\xi_o^2) - r - r + E\,[(y_\ell - y_o)(y_k - y_o)]$$

$$= E(\xi_o^2) - 2r + V\,x_k - 2\,V\,x_o + V\,x_o + r$$

$$= E(\xi_o^2) - r + k \cdot q \tag{B17}$$

$$E(\xi_\ell \xi_o) = E(\xi_o^2) + E(\xi_o(y_\ell - y_o)) = E(\xi_o^2) - r \tag{B18}$$

Using (B16), (B17), (B18) we can evaluate

$$V_{L+1} = E(\sum_o^L \xi_\ell)^2 \tag{B19}$$

$$V_{L+1} = \sum_o^L E\xi_\ell^2 + \sum_{\ell \neq k} E\,\xi_\ell \xi_k \tag{B20}$$

There are L+1 terms such that

$$\sum_{\ell=o}^{L} E(\xi_o^2) + \ell q = (L+1)\,E(\xi_o^2) + q\frac{L(L+1)}{2} \tag{B21}$$

There are 2L terms in second sum where ℓ=o or k=o

$$2LE(\xi_o^2) - r) \tag{B22}$$

There are L^2-L terms remaining in which there are 2(L-k) pairs involving ξ_k for k < ℓ Hence

$$\sum_{k=1}^{L} 2(L-k)\;(E(\xi_o^2) - r + kq)$$

$$= (E\xi_o^2 - r)(2L^2 - \frac{2L(L+1)}{2} + 2q\,[\frac{L^2(L+1)}{2} - \frac{(L+1)L(2L+1)}{6}]$$

$$= (E\xi_o^2 - r)(L^2 - L) + q\,\frac{L(L^2-1)}{3} \tag{B23}$$

Combining (B21), (B22), (B23)

$$V_{L+1} = (L+1)^2\,V_1 - L(L+1)\,r + \frac{(2L^2 + 3L + 1)\,L}{6}\,q \tag{B24}$$

(B24) $\Rightarrow$ Eqn (2.24) QED.

B.2 Linear Growth with Constant β

$$y_n = x_n + \eta_n$$

$$x_n = x_{n-1} + \beta + v_n$$

Expression (B6) is augmented by a term

$$[(B+1)\beta - (\frac{\beta}{B}\sum_{1}^{B-1} j + \beta)]$$ for which the expected square

$$\text{is } (\frac{B+1}{2}\beta)^2 = \delta_\beta \tag{B25}$$

(B25) + (B9) $\Rightarrow$ (3.27) QED

To find $V_{L,\beta}$ we proceed as in Section B.1

$$E(\xi_\ell^2) = E(x_\ell + \eta_\ell - x_o)^2 = E(x_o - \hat{x}_o + \ell\beta + \eta_\ell + \sum_{o}^{\ell-1} v_j)^2$$

$$= E(x_o - \hat{x}_o)^2 + r + \ell q + \ell^2\beta^2 + 2\ell\beta E(x_o - \hat{x}_o) \tag{B26}$$

First 3 terms can be expressed using (B16) and (B25) as

$$E(\xi_\ell^2)_{\beta=o} + \delta_\beta \text{ ; and } E(x_o - \hat{x}_o) = B\beta - \beta/B \sum_{o}^{B-1} j$$

$$E(\xi_\ell^2) = E(\xi_\ell^2)_{\beta=o} + \delta_\beta + \ell^2\beta^2 + 2\ell\beta(B\beta - \beta(B-1)(B)/2B)$$

$$= E(\xi_\ell^2)_{\beta=o} + \delta_\beta + \beta^2\ell(\ell+B+1) \tag{B27}$$

$$E(\xi_\ell \xi_o) = E[(x_o - \hat{x}_o + \ell\beta + \eta_\ell + \sum_o^{\ell-1} v_j)(x_o - \hat{x}_o + \eta_o)]$$

$$= E(\xi_\ell \xi_o)_{\beta=o} + \delta_\beta + \ell\beta E(x_o - \hat{x}_o)$$

$$\beta^2 \ell \frac{B+1}{2} \qquad \text{(B28)}$$

$$E(\xi_\ell \xi_k) = E(x_o - \hat{x}_o + \ell\beta + \eta_\ell + \sum v_j)(x_o - \hat{x}_o + k\beta + \eta_k + \sum^{k-1} v_j)$$

$$= E(\xi_\ell \xi_k)_{\beta=o} + \delta_\beta + \ell k\beta^2 + (\ell+k)\beta^2 \frac{B+1}{2} \qquad \text{(B29)}$$

(B29) is found by expansion analogously to (B28)

For a leadtime L+1, we have additional terms in (B19).

From (B27)

$$\text{i)}\ \beta^2 \sum_o^L \ell(\ell+B+1) = \left[\frac{(L+1)L(2L+L)}{6} + (B+1)\frac{L(L+1)}{2}\right]\beta^2$$

From (B28)

$$\text{ii)}\ \beta^2 \sum_o^L L\frac{B+1}{2} = \beta^2(B+1)\frac{L(L+1)}{2}$$

From (B29)

$$\text{iii)}\ \beta^2\, 2 \sum_{\ell=1}^{L} \sum_{k=1}^{\ell-1} (\ell k + \ell k \frac{B+1}{2})$$

$$= 2\beta^2 \sum_1^L \ell\frac{(\ell-1)\ell}{2} + \frac{B+1}{2}[\ell^2 + \frac{(\ell-1)\ell}{2}]$$

$$= \beta^2 \sum_1^L [\ell^3 - \ell^2 + (B+1)\frac{3\ell^2 - \ell}{2}]$$

$$= \beta^2 \frac{L^2(L+1)^2}{4} - \frac{L(L+1)(2L+L)}{6} + (B+1)\frac{(L+L)L(2L+1) - L(L+1)}{4}$$

$$\text{iv)}\quad (L+1)^2\delta_\beta$$

Combining i), ii), iii) and iv) we find additional terms to add to $V_{L+1,o}$,

$$\frac{\beta^2}{4} L^2 (L+1)^2 + (B+1)\, 2L(L+1)(L+2) + (L+1)^2\, \delta_\beta \qquad \text{(B30)}$$

(B30) with (B25) $\Rightarrow$ (3.28)

APPENDIX C

STATISTICAL PROCEDURES FOR OBTAINING k-FACTORS

As evidenced from equations (2.12), (2.26), (3.26) and (2.5) for r_n^2, q_n^2 constant; the process parameter $k = r^2/q^2$ is quite important as a parameter of the forecasting algorithms. One would like to know the "k-factor" for each individual time series (item*) being generated by the process models discussed in Chapters III and IV and forecast future periods' observables accordingly. It is usually not feasible or practical to obtain k-factors for each individual item; infeasible because there is often not enough data to use a portion for obtaining k-estimates and the remainder for forecast testing; impractical because for an inventory of many items one may be only able to retain, in the forecast system, parameters for classes of items.

During computer runs using MA algorithms with base B, squared errors can be averaged over the time horizon by each item i to provide statistical estimates $V_{1i}(B)$, $V_{Li}(B)$. Equations (2.23), (2.24) or (3.27), (3.28) can be manipulated to obtain estimates of k. Items are stratified by some classification scheme into cells and several methods for using the cell averages of V_1, V_L, k are posed.

All the methods below except e. assume a dynamic mean model, i.e., (2.23) and (2.24) are used to find k.

a. Determine $k_i = f\left(\frac{\hat{V}_{Li}(B)}{\hat{V}_{1i}(B)}\right)$ or $k_i = g\,(\hat{V}_{1i}(B),\ \hat{V}_{1i}(B'))$,

B, B', L in the equations being known.

Then average k_i over items in cell $\rightarrow \hat{k}$

b. Let $\hat{\hat{V}}_1(B)$ = average over items in cell of $\hat{V}_{1i}(B)$

Then $k = g\,(\hat{\hat{V}}_1(B),\ \hat{\hat{V}}_1(B')$

---

*We discuss the problem in the context of demand over time for an item.

c. Let $\hat{\hat{R}}$ = average ratio over items in cell of $\dfrac{\hat{V}_{Li}(B)}{\hat{V}_{1i}(B)}$

Then $k = f(\hat{\hat{R}})$

d. $\hat{k} = f\left(\dfrac{\hat{\hat{V}}_L(B)}{\hat{\hat{V}}_1(B)}\right)$

e. This used with log transformation, assuming a dynamic proportion model, where $\beta^2 = (-1/2\ q^2)^2$ in equations (3.27), (3.28).

For each cell, solve for $\hat{k}$, $\hat{q}^2$

$\hat{\hat{V}}_1(B)$ = function of $(\hat{k}, \hat{q}^2)$ eqn. (3.27)

$\hat{\hat{V}}_L(B)$ = function of $(\hat{k}, \hat{q}^2)$ eqn. (3.28)

<u>Discussion</u>:

The methods above were investigated in Orr [8] for the particular problems of obtaining k-factors by item class for four time series: Demand D, Demand per flying hour D/H, log D, and log D/H. The criterion for selection of a method and for selection of an item stratification scheme was evidence of a pattern of k values over item classes.

a. This method was not investigated. Variability in $\hat{V}_{Li}$, $\hat{V}_{1i}$ for individual items is large; k_i would be suspect. Also a few very large k_i's would dominate the cell average.

b. This method was tested in a simulation* against method c. and fared worse; a sample variance of $\hat{k}$ was larger and sample mean of $\hat{k}$ was further from a true k.

c. This procedure was adopted for obtaining $\hat{k}$ by cells for the D and D/H series. Reasonable k-patterns were obtained for several stratification schemes using real data for $\sim$ 10,000 items.

---

*A dynamic mean model with a given k generated observations moving averages with given bases were applied for forecasting and statistical estimates of MSE were used in equations to find $\hat{k}$. Monte Carlo replications gave statistics on the mean and variance of $\hat{k}$.

d. This procedure was tried using a stratification scheme which had done well in method c. It fared badly, yielding negative k's and no pattern.

e. This procedure was used in obtaining $\hat{k}$, $\hat{q}^2$ by cells for log D and log D/H series. Patterns were obtained for same strats (with method c.) used on D and D/H series.

DISTRIBUTION

COPIES	
1	Deputy Under Sec'y of the Army, ATTN: Office of Op Resch
	Headquarters, US Army Materiel Development & Readiness Command
1	DRCPA
1	DRCPA-S
1	DRCDMR
1	DRCPS
1	DRCMM
1	DRCMM-R
1	DRCMM-RS
1	DRCMM-S
1	DRCMM-M
1	DRCRE
1	Dep Chf of Staff for Logistics, ATTN: DALO-SMS, Pentagon, Wash., DC 20310
1	Asst Dep Chf of Staff for Logistics (Supply & Maintenance), Pentagon, Wash., DC 20310
2	Defense Logistics Studies Info Exchange, DRXMC-D
2	Commander, US Army Logistics Center, Ft. Lee, VA 23801
1	Commander, US Army Tank-Automotive Research & Development Cmd, ATTN: DRDTA-V, Warren, MI 48090
1	Commander, US Army Tank-Automotive Materiel Readiness Cmd, ATTN: DRSTA-S, Warren, MI 48090
1	Commander, USA Armament Command, Rock Island, Ill 61201
1	Commander, USA Electronics Command, Ft. Monmouth, NJ 07703
1	Commander, USA Missile Command, Redstone Arsenal, Ala 35809
1	Commander, USA Missile Command, ATTN: Mr. Ray Dotson, DRSMI-DA
1	Commander, USA Troop Support Command, 4300 Goodfellow Blvd., St. Louis, MO 63120
1	Commander, USA Aviation Systems Command, P.O. Box 209, St. Louis, MO 63120
2	Commander, Army Automated Logistics Mgt Systems Agency, Box 14505, St. Louis, MO 63168
1	Director, DARCOM Logistics Systems Support Agency, Letterkenny Army Depot, Chambersburg, PA 17201
1	Commander, Maintenance Mgt Center, Lexington-Blue Grass Army Depot, Lexington, KY 40507
1	Director, Army Management Engineering Training Agency, Rock Island Arsenal, Rock Island, Ill 61202
5	Commandant, US Army Logistics Management Center, Ft. Lee, VA 23801
1	Defense Supply Agency, DSAH-LO, Cameron Sta, Alexandria, VA 22314
4	Deputy Chief of Staff (I&L), HQ USMC-LMP-2, Wash., DC 20380
10	Defense Documentation Center, Cameron Sta., Alexandria, VA

1 Commander, US Air Force Logistics Command, WPAFB, Dayton, Ohio, ATTN: AFLC/XRS 45433

1 US Navy Fleet Materiel Support Office, Naval Support Depot, Mechanicsburg, PA 17055

1 Mr. James Prichard, Navy Supply Systems Cmd, Dept of US Navy, Wash., DC

1 George Washington University, Inst of Management Science & Engineering, 707 22nd St., N.W., Washington, DC

1 Naval Postgraduate School, ATTN: Dept of Opns Anal, Monterey, Calif 93940

1 Air Force Institute of Technology, ATTN: SLGQ, Head Quantitative Studies Dept., Dayton, Ohio 43433

2 The Army Library, Room 1A518, Pentagon, Wash., DC 20310

1 US Army Military Academy, West Point, NY

1 Logistics Management Institute, 4701 Sangamore Road, Wash.,DC 20016

1 General Research Corp., McLean, VA 22101

1 University of Florida, ATTN: Dept of Industrial & Systems Engineering, Gainesville, Fla

1 RAND Corp., 1700 Main St., Santa Monica, CA 90406

1 Office, Asst Sec'y of Defense (Inst & Logistics) ATTN: Mr. George Minter, Pentagon, Wash., DC 20310

1 Commander, AVSCOM, ATTN: Systems Analysis Directorate, P.O. Box 209, St. Louis, MO 63166

1 US Army Materiel Systems Analysis Activity, ATTN: AMXSY-CL, Aberdeen Proving Ground, MD 21005

1 Commander, US Army Logistics Center, ATTN: Studies Analysis Div., Concepts & Doctrine Directorate, Ft. Lee, VA 23801

1 Wallace M. Cohen, Asst Director Systems Analysis, FGMSD, General Accounting Ofc, Wash., DC 20548

1 ALOG Magazine, ATTN: Tom Johnson, USALMC, Ft. Lee, VA 23801

1 Commander, Air Force Logistics Cmd, ATTN: AFLC/AQMLE, WPAFB, Dayton, Ohio 45433

1 Operations Research Center, 3115 Etcheverry Hall, University of California, Berkeley, CA 94720

1 Major Keith Oppenmeer, HDQ Dept of the Army, (DASG-HCL-P), Wash., DC 20314

1 Dr. Jack Muckstadt, Dept of Industrial Engineering & Operations Research, Upson Hall, Cornell University, Ithaca, NY 14890

1 Prof Herbert P. Galliher, Dept of Industrial Engineering, University of Michigan, Ann Arbor, MI 48104

1 Mr. Ellwood Hurford, Scientific Advisor, Army Logistics Center, Ft. Lee, VA 23801

1 Commandant, USA Armor School, ATTN: MAJ Harold E. Burch, Leadership Dept, Ft. Knox, KY 40121

1 Prof Robert M. Stark, Dept of Stat & Computer Sciences, University of Delaware, Newark, DEL 19711

1 Prof E. Gerald Hurst, Jr., Dept of Decision Science, The Wharton School, University of Penna., Phila., PA 19174

1 Logistics Studies Office, DRXMC-LSO, ALMC, Ft. Lee, VA 23801

1 Procurement Research Office, DRXMC-PRO, ALMC, Ft. Lee, VA 23801

1 Commander, US Army Aviation Systems Cmd, ATTN: DRSAV-D, P.O. Box 209, St. Louis, MO 63166

1 Commander, US Army Electronics Cmd, ATTN: DRSEL-SA, Ft. Monmouth, NJ 07703

1 Commander, US Army Test & Evaluation Cmd, ATTN: DRSTE-SY, Aberdeen Proving Ground, MD 21005

1 Commander, US Army Armament Cmd, ATTN: DRSAR-SA, Rock Island, IL 61201

1 Commander, US Army Troop Support Command, ATTN: DRSTS-G, 4300 Goodfellow Blvd, St. Louis, MO 63120

1 Commander, US Army Mobility Equipment Research & Development Cmd, ATTN: DRXFB-O, Ft. Belvoir, VA 22060

1 Commander, US Army Natick Research & Development Cmd, ATTN: DRXNM-O, Natick, MA 01760

1 Commander, USA Missile Cmd, ATTN: DRSMI-C, Redstone Ars, AL 35809

1 Commander, USA Missile Cmd, ATTN: DRSMI-D, Redstone ARS, AL 35809

1 Commander, USA Armaments Researcy & Development Cmd, Implementation Task Force, ATTN: DRCSA-RP-SE (Mr. Larry Ostuni) Picatinny Arsenal, Dover, NJ 07801

1 Prof William S. Jewell, Dept of Ind Engrg & Operations Research, University of California, 4173 Etcheverry Hall, Berkeley, CA 94720